IBM Data Science- The Battle of Neighborhoods (Week 2

Chinese Restaurant Exploration

Problem Statement

A budding entrepreneur sees a trend in New York City for Chinese cuisine. He is interested in setting up shop in New York City. Being in the big city, he is unsure of where the best place would be for him to open up his new chinese restaurant. However, he understands the power of data in order to make a more informed decision. Therefore, utilising data, he would like to determine where the best place to open up a new chinese restaurant would be.

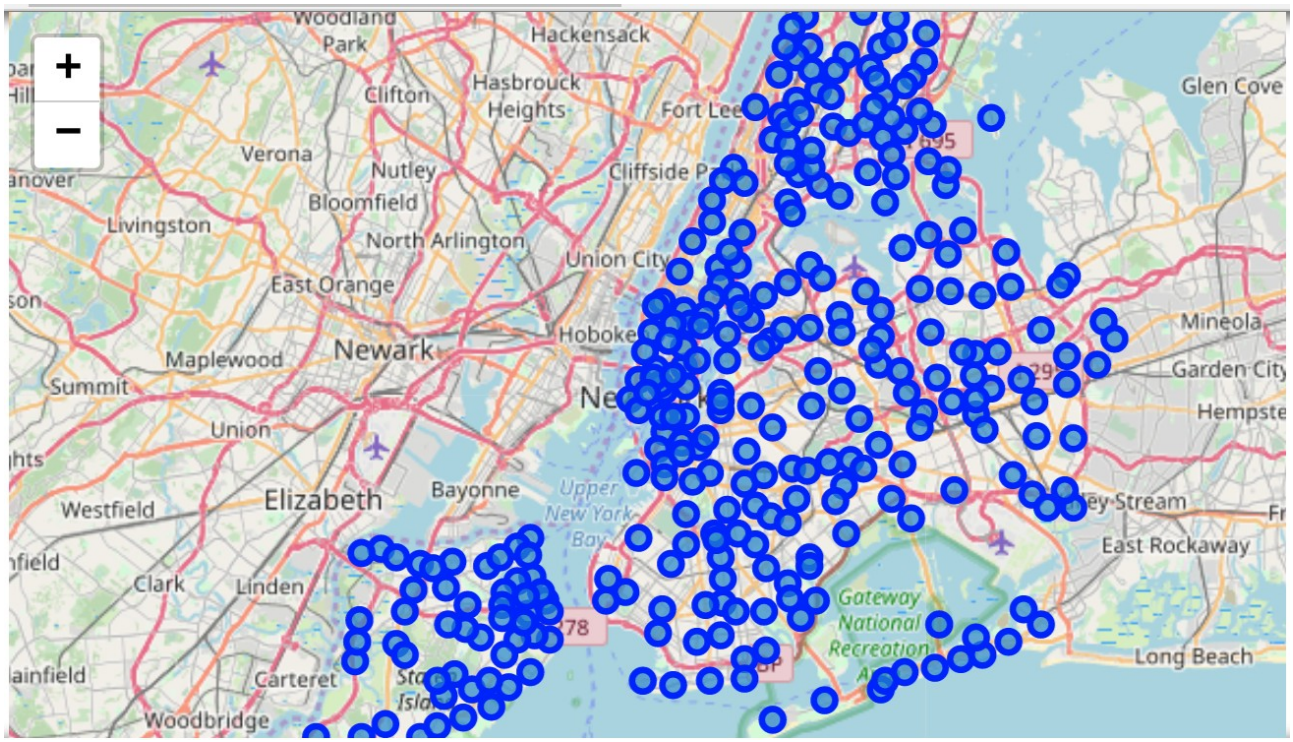
Data

In order to help this restaurateur solve his dilemma, data from the New York City neighbourhoods database will be used. Then, based on this dataset and the given lattitude and longitudes of the neighbourhoods, the Foursquare API will be queried in order to find out the distribution of Chinese restaurants in each neighbourhood. Then, based on the neighbourhood with the lowest density of Chinese restaurants available, it could be used to determine whether or not the new restaurateur should set up shop there.

Methodology

First, the dataset is downloaded and converted to a pandas dataframe.

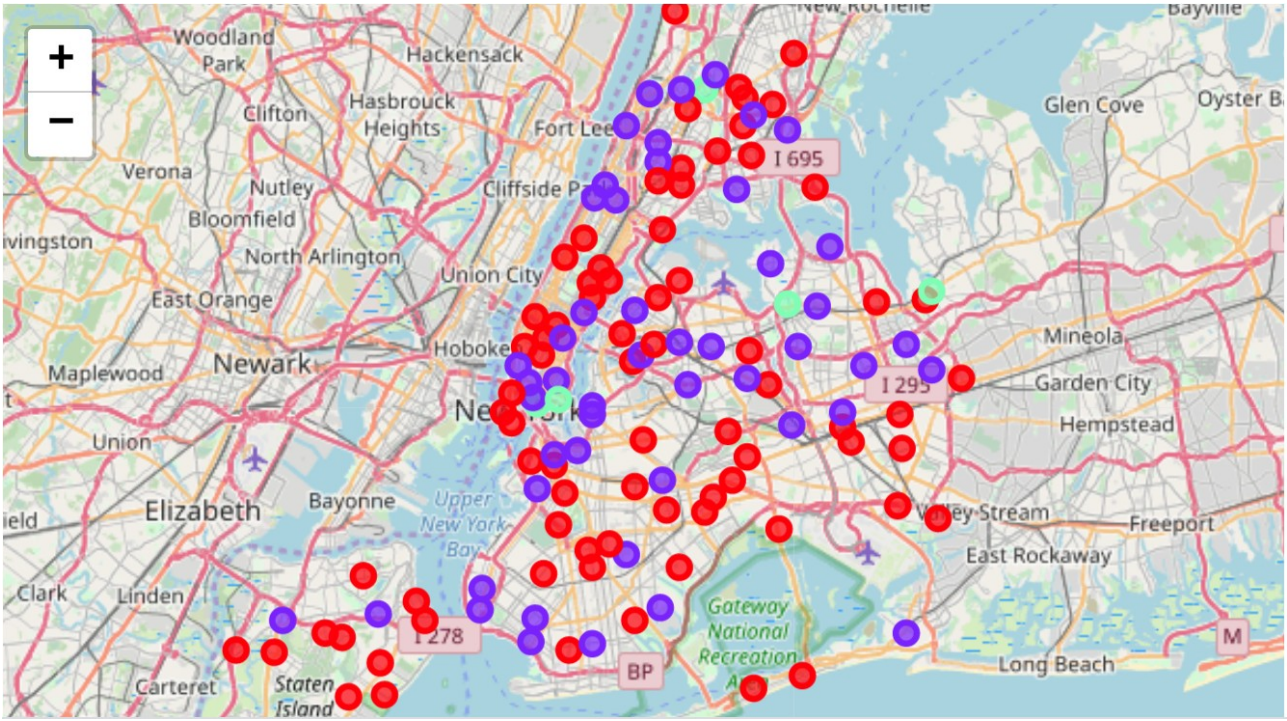
Next, some exploratory data analysis is performed on the imported dataset.



Then, the number of venues of each neighborhood is obtained, and the Chinese Restaurant venues are filtered out. Then, the sum of Chinese restaurants for each area is calculated.

After that, clustering is performed to determine the areas with the lowest amount of Chinese restaurants. The clustering is split into 3 clusters with Kmeans clustering.

Once the clusters are obtained, the clusters are plotted on the neighborhoods.



Results & Discussion

	Neighborhood	Chinese Restaurant	Borough	Latitude	Longitude
0	Allerton	1	Bronx	40.865788	-73.859319
1	Astoria	1	Queens	40.768509	-73.915654
3	Battery Park City	1	Manhattan	40.711932	-74.016869
5	Bayside	1	Queens	40.766041	-73.774274
9	Belle Harbor	1	Queens	40.576156	-73.854018
10	Bellerose	1	Queens	40.728573	-73.720128
12	Blissville	1	Queens	40.737251	-73.932442
13	Boerum Hill	1	Brooklyn	40.685683	-73.983748
14	Borough Park	1	Brooklyn	40.633131	-73.990498
15	Bronxdale	1	Bronx	40.852723	-73.861726
16	Brownsville	1	Brooklyn	40.663950	-73.910235
18	Bushwick	1	Brooklyn	40.698116	-73.925258
19	Canarsie	1	Brooklyn	40.635564	-73.902093
20	Carnegie Hill	1	Manhattan	40.782683	-73.953256
22	Chelsea	1	Manhattan	40.744035	-74.003116
22	Chelsea	1	Staten Island	40.594726	-74.189560
24	City Line	1	Brooklyn	40.678570	-73.867976
25	Claremont Village	1	Bronx	40.831428	-73.901199
26	Clifton	1	Staten Island	40.619178	-74.072642
27	Clinton	1	Manhattan	40.759101	-73.996119
29	Cobble Hill	1	Brooklyn	40.687920	-73.998561
32	Concourse Village	1	Bronx	40.824780	-73.915847
33	Corona	1	Queens	40.742382	-73.856825

Neighborhood	Chinese Restaurant	Borough	Latitude	Longitude	
34	Ditmas Park	1	Brooklyn	40.643675	-73.961013
35	Dongan Hills	1	Staten Island	40.588673	-74.096399
36	Douglaston	1	Queens	40.766846	-73.742498
39	East New York	1	Brooklyn	40.669926	-73.880699
41	Eastchester	1	Bronx	40.887556	-73.827806
42	Edgewater Park	1	Bronx	40.821986	-73.813885
44	Eltingville	1	Staten Island	40.542231	-74.164331
45	Erasmus	1	Brooklyn	40.646926	-73.948177
47	Financial District	1	Manhattan	40.707107	-74.010665
48	Flatbush	1	Brooklyn	40.636326	-73.958401
49	Flatiron	1	Manhattan	40.739673	-73.990947
51	Fordham	1	Bronx	40.860997	-73.896427
52	Forest Hills	1	Queens	40.725264	-73.844475
55	Glendale	1	Queens	40.702762	-73.870742
57	Gravesend	1	Brooklyn	40.595260	-73.973471
58	Great Kills	1	Staten Island	40.549480	-74.149324
61	Hollis	1	Queens	40.711243	-73.759250
63	Howard Beach	1	Queens	40.654225	-73.838138
65	Jamaica Center	1	Queens	40.704657	-73.796902
69	Lenox Hill	1	Manhattan	40.768113	-73.958860
72	Long Island City	1	Queens	40.750217	-73.939202
74	Manhattan Valley	1	Manhattan	40.797307	-73.964286
76	Manor Heights	1	Staten Island	40.601810	-74.120594
77	Marine Park	1	Brooklyn	40.609748	-73.931344
79	Midland Beach	1	Staten Island	40.573527	-74.093483
80	Midtown	1	Manhattan	40.754691	-73.981669
81	Midtown South	1	Manhattan	40.748510	-73.988713
83	Morrisania	1	Bronx	40.823592	-73.901506
86	New Dorp	1	Staten Island	40.572572	-74.116479
87	New Lots	1	Brooklyn	40.662744	-73.885118
88	New Springville	1	Staten Island	40.594252	-74.164960
89	North Riverdale	1	Bronx	40.908543	-73.904531
94	Olinville	1	Bronx	40.871371	-73.863324
95	Park Slope	1	Brooklyn	40.672321	-73.977050
96	Parkchester	1	Bronx	40.837938	-73.856003
98	Pelham Gardens	1	Bronx	40.862966	-73.841612
100	Port Morris	1	Bronx	40.801664	-73.913221
104	Rockaway Beach	1	Queens	40.582802	-73.822361
105	Rosedale	1	Queens	40.659816	-73.735261
106	Rossville	1	Staten Island	40.549404	-74.215729
107	Shore Acres	1	Staten Island	40.609719	-74.066678
109	South Jamaica	1	Queens	40.696911	-73.790426
111	Springfield Gardens	1	Queens	40.666230	-73.760421
112	St. Albans	1	Queens	40.694445	-73.758676
113	Steinway	1	Queens	40.775923	-73.902290
115	Sunnyside Gardens	1	Queens	40.745652	-73.918193
117	Tribeca	1	Manhattan	40.721522	-74.010683
118	Upper East Side	1	Manhattan	40.775639	-73.960508
119	Upper West Side	1	Manhattan	40.787658	-73.977059
121	Weeksville	1	Brooklyn	40.675040	-73.930531
122	West Brighton	1	Staten Island	40.631879	-74.107182

Neighborhood	Chinese Restaurant	Borough	Latitude	Longitude	
123	West Farms	1	Bronx	40.839475	-73.877745
125	Willowbrook	1	Staten Island	40.603707	-74.132084
126	Windsor Terrace	1	Brooklyn	40.656946	-73.980073
127	Woodhaven	1	Queens	40.689887	-73.858110
128	Woodrow	1	Staten Island	40.541968	-74.205246
130	Yorkville	1	Manhattan	40.775930	-73.947118

From the analysis conducted, it can be observed that cluster 1 contains the areas with the lowest count of Chinese Restaurants. Overall, it can be observed that New York City still has a lot of opportunity for new Chinese Restaurants. Other than Manhattan, Queens, and the Bronx area, the number of Chinese Restaurants is still relatively low. One of the boroughs with the lowest amount of Chinese Restaurants would be the Staten Island area.

Conclusion

In conclusion, analysis was done on the New York City Neighborhoods dataset, and a suitable new location to open up a Chinese restaurant was determined.